ChinaXiv合作期刊

知识管理论坛 ISSN 2095-5472 CN11-6036/C
Knowledge Management Forum

E-mail: kmf@mail.las.ac.cn
http://www.kmf.ac.cn

【学术探索】

基于多源异构数据挖掘的“红色记忆”知识图谱构建

◎ 郭嘉欣

华中师范大学信息管理学院 武汉 430079

摘要：[**目的 / 意义**] 中华民族与中国共产党人对真理追求的过程形成红色文化资源，对其进行知识组织和挖掘构建“红色记忆”，不仅能够提升民族自信与凝聚力，更是坚定文化自信的重要途径。针对红色文化资源所存在的分布广、来源多、类型杂、内容有限、组织程度低等问题，构建基于多源异构数据挖掘的“红色记忆”知识图谱，以充分利用红色文化资源。[**方法 / 过程**] 首先通过设计概念、关系及属性构建红色文化资源本体库，完成“红色记忆”的知识建模工作；其次通过多渠道采集红色文化资源，具体分析红色文化资源的构成和特点，针对这些多源异构数据进行实体、属性、关系识别采取；最后通过图数据库存储构建“红色记忆”知识图谱。[**结果 / 结论**] 通过构建“红色记忆”知识图谱，能够对多源异构的红色文化资源数据进行深层关系挖掘，提升红色文化资源的组织水平，为实现红色文化智能化服务奠定基础。

关键词：红色文化资源 知识图谱构建 知识建模

分类号：G250

引用格式：郭嘉欣 . 基于多源异构数据挖掘的“红色记忆”知识图谱构建 [J/OL]. 知识管理论坛 , 2020, 5(1): 59-68[引用日期]. http://www.kmf.ac.cn/p/200/.

1 引言

红色文化资源是中华民族与中国共产党人在对真理追寻的过程中形成的，这使得其历史发展的周期性较长，从而导致红色文化资源在开发和利用的过程中存在着分布广、来源多、类型杂、内容有限、组织程度低等问题，阻碍了用户对红色文化资源的深层次利用。2012 年，谷歌公司首先提出知识图谱的概念 [1]，意在从语义角度组织网络数据，构建大型知识库，进而提供智能搜索服务。国内外各公司和研究机构也纷纷开始构建知识图谱，如德国马普所的 YAGO[2]、谷歌的 Knowledge Vault[3]、复旦大学的 CN-DBpedia[4] 及清华大学的 XLore[5] 等。知识图谱作为一种重要的知识表示方式，逐渐成为各行各业从网络化向智能化转型升级的重要一环，具有广阔的发展前景 [3]。

红色文化资源作为中华优秀文化的重要

作者简介：郭嘉欣 (ORCID：0000-0002-7243-0291)，硕士研究生，E-mail：jxguo718@163.com。

收稿日期：2019-12-10 发表日期：2020-02-25 本文责任编辑：刘远颖

构成部分，蕴涵着十分丰富的革命和历史价值，是坚定文化自信的基础支撑[6]。受电子技术迅速发展的影响，许多地区提出了建立红色文资源数据库，如四川特色文化资源数据库[7]、西柏坡红色教育资源基础数据库[8]等，这在一定程度上使得红色文化资源的组织程度得到了提升，但也还仅仅停留在数据存储的阶段，其组织程度还不够高。知识图谱这一新的资源组织方式并没有在红色文化资源的研究利用中得到广泛的应用。因此，笔者通过采集结构各异、来源不同的红色文化资源数据，对其进行知识组织和挖掘，进而构建"红色记忆"知识图谱，提升红色文化资源组织程度，把红色文化资源以更直观、动态、关联的形式呈现给用户。

2 "红色记忆"知识图谱的构建流程

红色文化资源是中国共产党领导的革命和建设中所形成的崇高精神及其物质载体的总称[9]，它不仅存在于过去，而且发展于当下，其内涵将伴随历史进程和实践需要而不断深化。对红色文化资源进行组织和挖掘，可以重现蕴涵在其中的"红色记忆"。知识图谱本质上是结构化、语义化的知识库，它以图的结构表示现实世界中的实体、属性及其关联，其中图的节点代表实体，而实体之间存在的语义关联则用图中的边来描述[10]。构建知识图谱的方式主要有以下两种：自顶向下和自底向上[11]。自顶向下的方式是指事先细化概念及概念之间的关系，完成本体库设计，形成知识图谱的 Schema 层，然后将实体匹配填充到预定义好的本体 Schema 层中。自底向上的方法则是先从语料库或数据集中抽取出实体、属性和关系，并把同类型的实体重新进行组织，将其抽象为概念，最后构建得到 Schema 层。

笔者将综合应用自顶向下和自底向上这两种不同的方式来构建"红色记忆"知识图谱。首先，通过观察比较红色文化资源的各个数据源，确定"红色记忆"知识图谱所需要的具体数据，通过编写网络爬虫、手动采集等方式从红色图书、网站、开放数据集、百科等多种数据源中获取构建"红色记忆"知识图谱所需要的数据，其中，开放数据集是结构化数据的主要来源，百科是半结构化数据的来源，从红色图书和红色文化垂直站点获取的则是非结构化文本；其次，通过剖析红色文化资源数据的构成及特征来设计概念、关系及属性，运用工具 Protégé 构建红色文化资源本体库，从而完成"红色记忆"知识建模；然后，基于设计好的本体库，根据所获取的不同形式的数据采取不同的方法进行实体、关系、属性的抽取；最后，将识别得到的红色文化资源知识进行整合处理，并将其存入图数据库 Neo4j 中，通过 Neo4j 完成知识的可视化呈现，实现"红色记忆"知识图谱的构建，整体过程如图 1 所示。

3 基于"红色记忆"本体构建的知识建模

知识建模是知识图谱构建的一项重要任务，它是对知识进行逻辑化和体系化的过程。通过本体构建来进行知识建模能够充分描述知识图谱中所涉及到事物的属性及联系。本体作为一种抽象化的表示模型，可以清楚明了地定义和描述概念及概念之间的关系，确定知识图谱的数据形态，说明知识图谱中存在哪些数据，例如实体的类别、不同实体所拥有的属性、实体与实体之间的关联[12]。本体的构建过程较为复杂，为了确保规范性，构建本体时必须要遵循相应的原则。目前被广泛认同的本体建模规范是 T. R. Gruber 提出的 5 条准则：明确性、一致性、可扩展性、最小编码偏差和最小本体承诺[13]。对于本体构建方法而言，目前已有一些较为成熟的方式，如 IDEF-5 法、Methontology 法、七步法和基于叙词表构建本体法等，其中，七步法相比其他方法而言具有一定的通用性[14]，所以笔者选用七步法，并综合考虑红色文化资源自身的特点，构建"红色记忆"本体库。

知识管理论坛，2020（1）:59-68
DOI: 10.13266/j.issn.2095-5472.2020.006

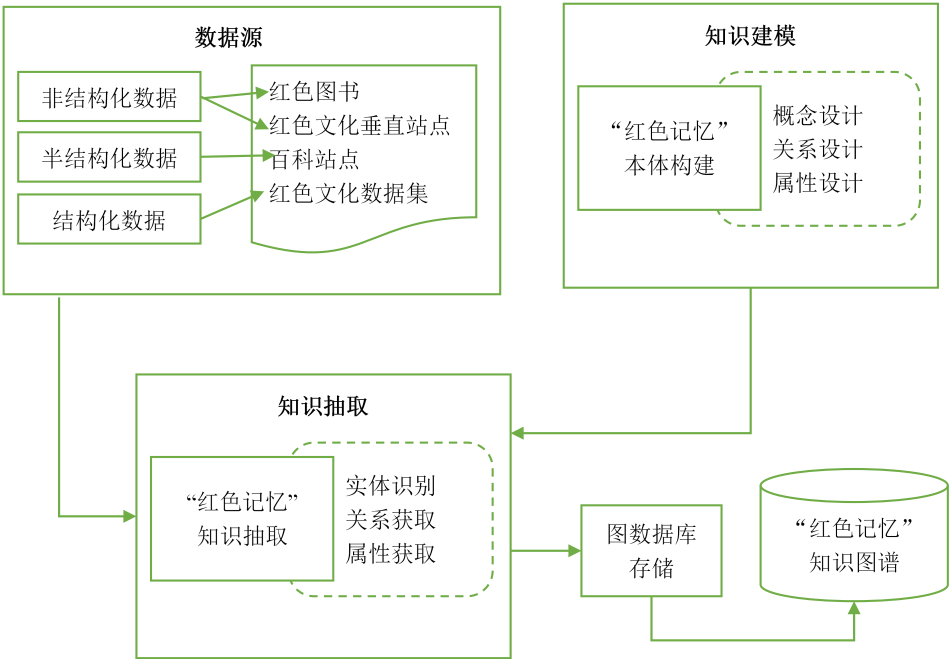


图 1 "红色记忆"知识图谱整体构建流程

作为一种特别的文化资源，红色文化资源不仅具有资源的属性也具有文化的属性，还具有二者深度融合所衍生出来的特殊属性[15]，这也导致了其分类标准的多样性。根据渠长根等[16]的归纳，目前学术界针对红色文化资源所采用的最基本的分类法是将其划分为物质和精神两大类，除此之外，有的学者将红色文化资源划分为动态和静态两种类型，或是根据一般、特殊的两分法来对红色文化资源进行分类。在实际的研究中，除了将红色文化资源按照简单的二分法标准来划分外，通常还会根据不同的学科需要来进行进一步的调整划分，张泰城[17]依据"以主题分类为主、兼顾学科的原则"，并遵循中文的语言习惯将红色文化资源划分为红色旧址、器物、文献、人物、事件、文艺、建筑、精神、研究、创作 10 个大类；张克伟[18]按照国家旅游资源的分类方法首先把红色文化资源细分为三大主类：遗址遗迹、建筑和设施、人文活动，再将其细分为 10 种基本类型，其中遗址遗迹包含历史事件的发生地、军事遗址与古战场两类，建筑和设施分为文化活动场所、展示演示的场馆、碑碣（林）、名人故居和历史纪念建筑、陵区陵园 5 类，人文活动包含人物、事件和文艺作品 3 类。

构建"红色记忆"本体库通常需要对概念、属性及关系等多个方面进行设计考量。对于"红色记忆"来说，其核心是人，因此首先确定的是"人物"这一重要概念，与之密切联系的必然是人物所经历或参与的事件，因此也加入"事件"概念。根据"人物"和"事件"这两个主题概念对"红色记忆"相关的信息进行浏览，发现人物所加入的组织与人物和事件的联系也非常密切，故将"组织"加入本体列表。除此之外，人物故居、纪念馆、陵园等信息也是比较重要的概念，而这些信息都可以看作是建筑，因此，新增"建筑"这一概念。针对"红色记忆"，其所具有的文化属性也必然会涉及到红色文学艺术作品，所以增加"资源"这一概念。

这 5 个概念确定之后，参考前文提到的分类标准以及实际搜集到的数据来辅助划分子概念。其中，人物作为独立概念不再进行划分；由于搜集到的事件相关数据基本为会议和战争两类，所以将事件划分为会议、战争及其他 3 类，同样地将组织划分为学校、军团、政党和其他，将建筑分为名人故居、纪念馆、纪念碑、纪念塔、遗址（旧址）、陵园、陵墓；资源则按照载体形态的不同分为书籍、电影、画作、诗词和歌曲。综合考虑以上几个概念，发现事件、建筑、组织的细分概念存在一些模糊的边界问题难以确定，并且直接使用子类概念进行构建会降低本体的可扩展性，所以将事件、建筑、组织的子类概念取消，转而新增“类型”这一概念，并将类型划分为事件类型、建筑类型和组织类型 3 类，并在各类型中添加“其他”这一选项，从而保证了所构建本体的全面性、准确性和可扩展性。

综上所述，“红色记忆”本体中的概念主要分为以下 6 个类别：建筑（Architecture）、事件(Event)、类型(Genre)、组织(Organization)、人物（Person）、资源（Resource），其中类型与资源两个概念下又划分了多个子概念，类别分为建筑类型、事件类型和组织类型。对每个类别的数据进行分析发现每个概念具有的特征不同，故根据不同类别的特征对属性进行定义，这里选取了“红色记忆”本体模型部分概念和属性进行展示，如表 1 所示：

表 1 “红色记忆”本体模型部分概念和属性

序号	概念	子概念	属性
1	建筑（Architecture）	建筑	建筑ID、名称、所在地、图片、类别、描述
2	事件（Event）	事件	事件ID、名称、起始时间、结束时间、发生地、参与人员、描述
3	类型（Genre）	组织类型	组织类型ID、组织具体类型
		建筑类型	建筑类型ID、建筑具体类型
		事件类型	事件类型ID、事件具体类型
4	组织（Organization）	组织	组织ID、组织名称
5	人物（Person）	人物	人物ID、姓名、出生时间、死亡时间、性别、职位、别名、国籍、民族、出生地、所著作品、描述
6	资源（Resource）	图书	图书ID、题名、出版时间、责任者、出版社、ISBN、描述
		诗词	诗词ID、题目、作者、具体内容、描述
		电影	电影ID、名称、类型、上映时间、参演演员、导演、编剧、描述

上述所设计的本体库中，概念和子概念之间是上下位关系，子概念具有不同的属性，子概念所包含的实体和实体之间则存在不同的语义关联，如人物与人物之间存在“配偶”、“子女”等多种关系，建筑与人物 / 事件之间存在“纪念”关系，基于前文所述概念设计，最终确定“红色记忆”中所涉及的部分关系。通过本体构建工具 Protégé 添加“红色记忆”定义好的概念及关系，完成“红色记忆”知识建模，设计的部分本体概念如图 2 所示：

表 2 “红色记忆”部分关系

关系名称	关系说明
配偶（hasMate）	人物与人物之间
子女（hasChild/isChildOf）	人物与人物之间
纪念（commemorateFor）	人物/事件与建筑
成员（hasMember）	组织与人物
事件类型（hasEventGenre）	事件与事件类型
参与（particaipateIn）	人物与事件

知识管理论坛，2020（1）:59–68
DOI：10.13266/j.issn.2095–5472.2020.006

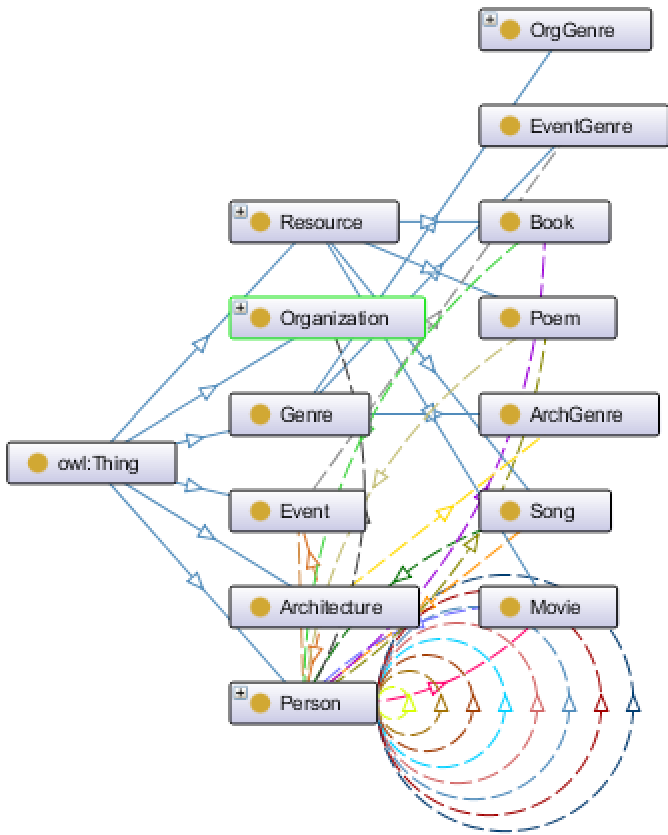


图 2 “红色记忆”部分概念本体

4 “红色记忆”数据源与知识获取

红色文化资源见证了我们党从成立之初到逐渐发展壮大的整个过程 [19]，其历史发展的周期较长，所以其资源采集、处理和存储方式也不尽相同，这也使得与红色文化资源有关的数据也呈现出明显的多源异构性。全国各地的图书馆、档案馆、博物馆及各类纪念馆、陈列馆、红色旅游景点等都是获取红色文化资源的来源，除此外，大数据时代的到来也使得各种 Web 资源变成获取红色文化资源的重要来源。所以，从这些来源采集到的结构化数据、半结构化数据和非结构化数据就是构建“红色记忆”知识图谱的数据基础。

结构化数据能够用数字或文字来描述或表达，具有相同的层次或网络结构，通常存储在关系型数据库中。“红色记忆”的结构化数据主要来源于开放数据集，具体方法是利用 API 接口将数据下载到本地并存储为关系型数据。“红色记忆”知识图谱的构建便是基于结构化数据，并搜集其他不同来源、不同结构的数据进行补充。

非结构化数据通常是利用自然语言形式保存的文本资源 [20]，是最丰富的知识来源，在红色文化网页、红色旅游网页、图书等非结构化的数据源中均存在大量文本。实体识别作为自然语言文本处理的基础 [21]，是知识图谱构建的重要步骤。实体识别即命名实体识别，是指从语料中抽取出具有特定含义的命名性指称项，如人名、地名及机构名等 [22]。对“红色记忆”知识图谱而言，要识别的实体即是在模式层的“红色记忆”本体模型中所定义的概念。对于实体识别，目前最常用的方法是通过机器学习

来实现，可以利用网络爬虫等相关工具从网页中获取“红色记忆”语料，再利用分词工具对语料进行分词、标注等预处理工作，之后将标注好的语料进行词向量转换。最后选取训练集语料，并通过机器学习训练出抽取模型 [23]，利用实体识别模型来从文本中提取出“红色记忆”的实体。实体识别完成后，可继续进行属性获取。

“红色记忆”知识图谱实体属性获取的来源是各类百科网站词条的 infobox，infbox 中的信息通常为半结构化数据，这些数据具有较高的一致性和完整性，这使得在获取“红色记忆”中人物信息时，只需利用爬虫爬取百科词条中相应的 infobox 标签即可获取关于人物的一些基本信息。例如图 3 展示的是“杨至成”这一人物词条的 360 百科的 infobox 信息，选取其中的中文名称、外文名称、别名、国籍 4 个属性，通过浏览网页源代码可以得到这些属性的信息（见图 4）。通过解析网页源代码，发现根据“class”标签找到人物所对应的属性，那么可以利用 python 的 BeautifulSoup 库来对 html 元素进行操作，从而获取“杨至成”的属性信息，得到 < 实体，属性，属性值 > 三元组。

基本信息

中文名称	杨至成	出生日期	1903年11月30日
外文名称	Zhicheng Yang	逝世日期	1967年2月3日
别名	杨至成	职业	军人
国籍	中华人民共和国	毕业院校	黄埔军校
民族	侗族	信仰	中国共产党
出生地	贵州省三穗县	主要成就	1955年被授予上将军衔 中南军区第一副参谋长

图 3 人物词条 infobox

```
▼<div class="card-list-box">
  ▼<ul class="cardlist">
    ▼<li>
      ▼<div class="cardlist-con">
          <p class="cardlist-name" title="中文名称">中文名称</p>
          <p class="cardlist-value" title="杨至成">杨至成</p>
        </div>
      </li>
    ▼<li>
      ▼<div class="cardlist-con">
          <p class="cardlist-name" title="外文名称">外文名称</p>
          <p class="cardlist-value" title="Zhicheng Yang">Zhicheng Yang</p>
        </div>
      </li>
    ▼<li>
      ▼<div class="cardlist-con">
          <p class="cardlist-name" title="别名">别名</p>
          <p class="cardlist-value" title=" 杨至成 "> 杨至成 </p>
        </div>
      </li>
    ▼<li>
      ▼<div class="cardlist-con">
          <p class="cardlist-name" title="国籍">国籍</p>
          <p class="cardlist-value" title=" 中华人民共和国 "> 中华人民共和国 </p>
        </div>
      </li>
```

图 4 网页源码

实体间关系的识别抽取则与实体识别的原理类似，再获取得到“红色记忆”实体后，结合所获得的“红色记忆”实体，选取含实体对象较多的语句，对其进行实体关系的抽取。通过对实体、属性、关系的识别抽取，最终获取到构建“红色记忆”

知识管理论坛，2020（1）:59–68
DOI：10.13266/j.issn.2095–5472.2020.006

知识图谱所需要的实体、属性和关系。最后，把从不同来源获取的数据进行整理归类，并将其存储在关系数据库中，部分数据示例如图 5 所示：

ArchID	nameS	address	place	firstImg	type	desc
1	杨至成故居	贵州省三穗县	(Null)	(Null)	(Null)	同时于2013年对杨至诚将军故居进行维修，故居位
2	周逸群烈士故	共同路12号	贵州省铜仁	(Null)	省政级爱	周逸群烈士故居原街名大公馆，1984年于国庆修复
3	王伯群上海故	上海市长宁区	上海市长宁	https://baike	省级重点	王伯群上海故居位于长宁区愚园路1136弄31号，该
4	王伯群贵阳故	贵阳市南明区	贵阳市南明	https://baike	省级文物	王伯群故居是贵阳历史上第一栋仿西式建筑， 时光
5	兴义王伯群故	贵州省兴义市	兴义市	https://baike	黔西南州	王伯群故居位于兴义市下五屯街道办事处景家屯，
6	总统府	江苏省南京市	南京市玄武	https://baike	全国重点	南京总统府，位于江苏省南京市玄武区长江路292号
7	段祺瑞执政府	北京市东城区	北京市东城	https://baike	中国人民	段祺瑞执政府，原为九贝子府，清康熙皇九子胤禟府
8	思群堂	上海市普陀区	上海市普陀	http://image	华东师范	华东师范大学前身学校之一大夏大学的师生1946年
9	龙大道烈士故	锦屏县茅坪镇	锦屏县	https://gss2.	省级文物	龙大道烈士故居位于锦屏县茅坪镇上寨村，距县城8
10	钱壮飞故居	嘉善路117弄	上海市	http://s13.si	上海市第	《徐汇区志》大事记中写道，1928年中共党员钱壮
11	邓恩铭故居	荔波县城北向	荔波县	https://gss3.	贵州省文	邓恩铭烈士故居为一栋四排三间土木结构的普通居民
12	杨至成将军纪	三穗县八弓镇	(Null)	(Null)	贵州省第一	杨至成将军纪念馆为中西结合式，有正厅和两厢。
13	周逸群陈列馆	中南门古城中	铜仁市碧江	(Null)	(Null)	周逸群烈士故居陈列馆，位于贵州碧江区中南门古城
14	遵义会议纪念	贵州省遵义市	遵义市	(Null)	国家一级	(Null)
15	四渡赤水纪念	贵州省习水县	遵义市	(Null)	国家二级	四渡赤水纪念馆包括四渡赤水纪念馆主馆、中国女
16	周逸群墓	湖南省	岳阳市君山	(Null)	(Null)	周逸群墓位于湖南省岳阳市君山区境内，周逸群烈士
17	中山陵	江苏省南京市	南京市玄武	https://baike	全国重点	中山陵是中国近代伟大的民主革命先行者孙中山先生
18	钱壮飞同志	贵州省毕节市	毕节市	https://gss1.	国家安全	(Null)
19	龙大道烈士陵	贵州省黔东南	黔东南苗族	https://gss3.	省级爱国	锦屏县把龙大道烈士陵园修葺一新，新竖龙大道烈士

图 5 “红色记忆”部分数据示例

5 “红色记忆”知识存储

目前，知识图谱的存储工作主要是通过图数据库完成的。通过图数据库存储知识图谱，能够实现图数据的可视化，并能通过图数据库所提供的各种工具对知识图谱进行集成管理，能高效迅速地满足用户的各类需求。当前，Neo4j 以其优良的性能和简单的操作等优点，在各种图数据库中使用最为广泛。笔者将“红色记忆”知识图谱存储在 Neo4j 中，Neo4j 中的标签代表“红色记忆”中的概念，节点代表了“红色记忆”中的实体，而边则描述的是关系。Neo4j 通过执行 Cypher 命令能够管理和操作知识图谱中的数据。由于 Cypher 命令提供批量导入 CSV 格式数据的 Load 语句，所以将关系型数据库中的“红色记忆”知识转化为 CSV 格式的文件进行存储，并按照以下语句批量导入。

批量导入概念 / 实体（以导入“建筑”为例）：

LOAD CSV WITH HEADERS FROM “file:///Architecture.csv” AS line

MERGE(p:Architecture{ArchID:line.ArchID,nameS:line.nameS,address:line.address,place:line.place,firstImg:line.firstImg,type:line.type,desc:line.desc})

批量导入关系（以导入人物与事件之间的关系“ParticiPateIn”为例）：

LOAD CSV WITH HEADERS FROM “file:///PersonToEvent.csv” AS line

Match(from:Person{PersonID:line.PersonID}),(to:Event{EventID:line.EventID})

merge(from)-[r:participateIn{PersonID:line.PersonID,EventID:line.EventID}]->(to)）

将存储在关系型数据库中的“红色记忆”知识批量导入 Neo4j 后形成“红色记忆”知识图谱，结果如图 6 所示，蓝色的圆点表示人物，

绿色的圆点表示组织，红色圆点表示建筑，棕色圆点表示红色资源，橙色原点代表事件，通过箭头指示它们之间的关系。由于知识图谱所具有的开放互联的特性，后续还可运用 Cypher 命令增加新的数据 [24]，形成大规模“红色记忆”知识图谱，从而实现红色文化智能搜索、知识问答、知识推理等应用，为实现红色文化资源的智能化服务奠定基础。



图 6 “红色记忆”知识图谱

6 结语

将知识图谱这一新的组织技术应用于红色文化资源的开发研究，是红色文化资源学科发展的必然抉择，也是数字化、智能化的时代要求。笔者通过定义概念、属性、关系设计了“红色记忆”本体库，完成“红色记忆”知识建模，并从结构不同、来源各异的红色文化数据源获取数据，基于这些数据进行命名实体的识别、关系及属性抽取来获取知识，进而得到“红色记忆”三元组，并将其存储于Neo4j中，构建了“红色记忆”知识图谱，从而更进一步地提升红色文化资源的组织程度，将红色文化资源以更直观、更现代的方式呈现出来，使得分布于各处的碎片化红色文化资源得到了重组 [25]，重现了蕴涵在书籍、歌曲、遗址中的“红色记忆”。在后

知识管理论坛，2020（1）:59–68
DOI：10.13266/j.issn.2095–5472.2020.006

续的工作中，笔者将进一步对“红色记忆”知识图谱的智能问答、知识推理等应用进行研究，满足用户对于红色文化的智能化服务的需求，更大程度上发挥红色文化资源中所蕴含的价值。

参考文献：

[1] SINGHA A. Introducing the knowledge graph: things, not strings[EB/OL]. [2019-04-10]. http://googleblog.blogspot.co.uk/2012/05/introducing-knowledge-graph-things-not.html.
[2] SUCHANEK F M, KASNECI G, WEIKUM G. Yago: a core of semantic knowledge[C]//Proceedings of the 16th international conference on World Wide Web. New York: ACM, 2007: 697-706.
[3] DONG X, GABRILOVICH E, HEITZ G, et al. Knowledge vault: a web-scale approach to probabilistic knowledge fusion[C] //International conference on knowledge discovery and data mining. New York: ACM, 2014: 601-610.
[4] XU B, XU Y, LIANG J, et al. CN-DBpedia: a never-ending Chinese knowledge extraction system[C]// International conference on industrial, engineering and other applications of applied intelligent systems. Berlin: Springer, 2017: 428-438.
[5] WANG Z, LI J, WANG Z, et al. XLore: a large-scale English-Chinese bilingual knowledge graph[C]// International semantic Web conference. New York: ACM, 2013: 121-124.
[6] 胡果 , 张荣秀 . 中华红色文化的主要特质及时代价值 [J]. 山西广播电视大学学报 , 2017(1): 103-105.
[7] 王茂春 . 特色文化资源与高新技术融合的路径探索 [J]. 中华文化论坛 , 2015(6): 128-133.
[8] 王玉平 , 张同乐 , 张志永 . 西柏坡红色文化资源数据库建设热议 [J]. 河北师范大学学报 (哲学社会科学版), 2014, 37(1): 140-145.
[9] 李实 . 准确认识“红色资源”的丰富内涵 [J]. 政工学刊 , 2005(12): 23.
[10] 漆桂林 , 高桓 , 吴天星 . 知识图谱研究进展 [J]. 情报工程 , 2017, 3(1): 4-25.
[11] 刘峤 , 李杨 , 段宏 , 等 . 知识图谱构建技术综述 [J]. 计算机研究与发展 , 2016, 53(3): 582-600.
[12] 马灿 . 面向“智慧法院”的知识图谱构建方法与研究 [D]. 贵州 : 贵州大学 , 2019.
[13] GRUBER T R. Toward principles for the design of ontologies used for knowledge sharing?[J]. International journal of human-computer studies, 1995, 43(5/6): 907-928.
[14] 岳丽欣 , 刘文云 . 国内外领域本体构建方法的比较研究 [J]. 情报理论与实践 , 2016, 39(8): 119-125.
[15] 张泰城 . 论红色文化资源 [J]. 红色文化资源研究 , 2015, 1(1): 1-11.
[16] 渠长根 , 闻洁璐 . 红色文化资源研究综述 [J]. 浙江理工大学学报 (社会科学版), 2019, 42(2): 179-187.
[17] 张泰城 . 论红色文化资源的分类 [J]. 中国井冈山干部学院学报 , 2017, 10(4): 137-144.
[18] 张克伟 . 沂蒙红色文化资源产业化研究 [D]. 济南 : 山东大学 , 2010.
[19] 许庆领 . 人文地理信息整合及可视化关键技术研究 [D]. 阜新 : 辽宁工程技术大学 , 2012.
[20] 郭文龙 . 中医方剂知识图谱构建研究与实现 [D]. 兰州 : 兰州大学 , 2019.
[21] 张晓艳 , 王挺 , 陈火旺 . 命名实体识别研究 [J]. 计算机科学 , 2005(4): 44-48.
[22] 王良萸 . 基于 web 数据的碳交易领域知识图谱构建研究 [D]. 马鞍山 : 安徽工业大学 , 2018.
[23] 蒋秉川 , 万刚 , 许剑 , 等 . 多源异构数据的大规模地理知识图谱构建 [J]. 测绘学报 , 2018, 47(8): 1051-1061.
[24] 吴雪峰 , 赵志凯 , 王莉 , 等 . 煤矿巷道支护领域知识图谱构建 [J]. 工矿自动化 , 2019, 45(6): 42-46.

KMF www.kmf.ac.cn
知识管理论坛
2020 年 第 1 期(总第 25 期)

Construction of "Red Memory" Knowledge Graph Based on Multi-source Heterogeneous Data Mining

Guo Jiaxin
Information Management School of Central China Normal University, Wuhan 430079

Abstract: [Purpose/significance] Red cultural resources are produced in the process of the Chinese nation and the Chinese Communists' pursuit of truth. Constructing "red memory" by organizing and mining knowledge of red cultural resources can not only enhance national self-confidence and cohesiveness, but also be an important part of cultural self-confidence. There may be many problems when using red cultural resources, such as wide distribution, multiple sources and types, limited content and low degree of organization. In order to make full use of red cultural resources, this paper constructs a "red memory" knowledge graph based on multi-source heterogeneous data. **[Method/process]** Firstly, this paper constructed a red cultural resource ontology library for knowledge modeling of "red memory". Secondly, it analyzed the composition and characteristics of red cultural resources collected through multiple channels and extract entities, attributes, relationships. Finally, the "red memory" knowledge graph was constructed through knowledge fusion and storage. **[Result/conclusion]** By constructing the "red memory" knowledge graph, it is possible to mine deep relationship on multi-source heterogeneous red cultural resource data, improve the organization degree of red cultural resources, and realize of intelligent services of red cultural resources.

Keywords: red cultural resources　knowledge graph construction　knowledge modeling